心理科学进展 2021, Vol. 29, No. 11, 2083–2090
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2021.02083

# 道德直觉合乎道义却不客观

袁晓劲 刘 昌

(南京师范大学心理学院, 南京 210097)

**摘 要** 面对道德困境, 道德直觉倾向于促使人们做出道义论的判断。但是, 道德直觉易受情绪因素影响, 具有较强的主观性。道德直觉警惕有意图地使用个人力量造成的伤害, 却会接受由非个人力量或连带作用引发的伤害。“模块近视假说”认为, 大脑中存在一个预警系统, 能快速地对主动伤害的想法发出情绪警报。但该系统的审查机制是一种简单的“单通道”加工, 这种加工局限使连带作用造成的伤害避开了审查机制的监控。道德直觉的不客观提示, 面对现实生活中的道德争议, 不应该仅听凭直觉作为行动的依据。

**关键词** 道德困境, 道德直觉, 情绪, 模块近视假说
**分类号** B849: C91



## 1 引言

道德问题是人类思想史上一个久经不息的讨论话题。当代研究者提出了一个“电车问题”(Trolley problem)思维实验, 用以探讨道德伦理问题(Foot, 1967; Thomson, 1976)。此举开启了道德研究领域的新视角, 引发了学界的兴趣, 催生了大量的研究参与其中。电车问题成为当代道德心理学的研究焦点之一(Greene, 2016a)。原始的电车问题也被称为开关困境(switch case), 该困境描述的是一列刹车失控的电车, 驾驶员面临着一个道德两难抉择: 是继续前行, 撞死主道上的五名工人, 还是扳动转向开关, 撞死支道上的一名工人? 扳动开关, “牺牲一个人挽救五个人”被认为是一种功利主义的选择。不扳动开关则被认为是一种道义论的选择(Greene, 2013)。

研究者从开关困境中衍生出其它道德困境, 用以探讨人们如何作出道德决策。其中一个称为人行天桥困境(footbridge dilemma), 该困境与开关困境类似, 不同之处在于, 挽救五个人的唯一方式是把天桥上的一个人推下铁轨挡停电车。研究中发现, 在开关困境中, 扳动转向开关, “以一换五”的选择占多数。但是, 在人行天桥困境中, 把人推下铁轨, “以一换五”的选择却占少数(Greene et al., 2001)。最近对来自 42 个国家的 7 万多名被试的一项重复性研究结果表明, 这一现象具有跨文化的普遍性(Awad et al., 2019)。“以一换五”的做法在开关困境里是可接受的, 但在人行天桥困境里就变成了不可接受, 这一结果引出了一个重要的疑问: 面对相似的道德问题, 人们为什么会表现出截然不同的态度?

近年来, 随着自动驾驶汽车的出现, 自动驾驶中的伦理难题成为了新时代的“电车问题”, 确切地出现在人们的日常生活中。自动驾驶同样面对着相似的困境: 面临事故时, 自动驾驶程序应该优先保护乘客还是行人? 或者是根据总体伤害程度决定(Bonnefon et al., 2016; Faulhaber et al., 2018; Maxmen, 2018)? 这些问题的解答均依赖于对人们如何做出道德判断的了解。文章从梳理“电车问题”的相关研究中, 探讨人们面对道德两难困境时的心理加工机制, 重点考察了情感因素在道德判断中所起的重要作用, 并引入了“模块近视假说”进行了解释。

## 2 情感引发合乎道义的直觉

研究者们采用功能性磁共振成像(fMRI)技术开展了一系列研究, 探讨道德困境决策的脑机制。他们设计了一系列道德困境, 一类与开关困境类似, 另一类与人行天桥困境类似, 让被试对

收稿日期: 2020-06-20
通信作者: 刘昌, E-mail:liuchang@njnu.edu.cn

这些道德困境进行判断。结果显示，与开关困境类似的道德判断更多激活了背外侧前额叶皮层(DLPFC)，而与人行天桥困境类似的道德判断更多激活了腹内侧前额叶皮层(VMPFC)及杏仁核(Amygdala)区域。另外，当面对与人行天桥困境类似的情境，如果被试做出符合功利主义的选择，DLPFC会产生更大程度的激活(Greene et al., 2004; Greene et al., 2009)。这些研究结果表明，人行天桥困境的判断与VMPFC及杏仁核的激活相关。开关困境的判断与DLPFC的激活相关。而且，DLPFC的激活与在人行天桥困境中做符合功利主义的选择也存在相关。

上述脑成像研究的一个局限是结果仅能反映相关性，无法确定是大脑的激活导致了人们的判断，还是大脑激活仅仅与人们的判断相关。神经病理学的研究对此提供了解答。在Mendez等人(2005)的研究中，选取了腹内侧前额叶皮层受损(Frontotemporal dementia, FTD)的病人、患有阿尔茨海默症(Alzheimer)的病人与健康对照组进行道德判断研究。结果表明，在开关困境中，赞同扳动开关的人占多数，并且三组之间的赞同率没有显著差异。在人行天桥困境中，阿尔茨海默症病人组与健康对照组赞同把人推下天桥的人占少数，且两组之间的差异不显著。但是，超过半数的额颞叶受损病人赞同把人推下天桥，赞同率远远高于另外两组。由此结果可以推断，情绪反应是导致人们在人行天桥中不赞同把人推下天桥的原因。其他研究者采用相同的逻辑相继进行探讨，结论均与Mendez等人的研究一致(Elisa et al., 2007; Koenigs et al., 2007; Young & Koenigs, 2007)。Crockett等人(2010)的研究中发现，让被试服用一种能加剧杏仁核和腹内侧前额叶皮层情绪反应的药物(Citalopram)后，被试在人行天桥类似困境中更不容易做出功利主义的判断。Perkins等人(2013)的研究发现，让被试服用一种会降低杏仁核和腹内侧前额叶皮层情绪反应药物(lorazepam)后，被试在道德困境中做出更多功利主义的判断。其他类似的研究也指向同样的结论：情绪反应使人们拒绝将天桥上的人推下铁轨(Trémolière et al., 2012; Gawronski et al., 2017; Gawronski et al., 2018)。总之，大量的研究证据表明，腹内侧前额叶皮层和杏仁核引起的情绪反应，导致人们在人行天桥类似的道德困境中更少做出功利主义的判断。

Greene等人(2001, 2004)整合了认知和情感两种因素在道德决策中的作用，提出了道德判断的双加工模型(dual-process model)。该模型认为，道德两难场景会引发两种分离的，通常是互相竞争的加工过程。一种是直觉反应，具有快速、自动化，通常包含情绪性信息的特点，另一种是推理反应，具有加工缓慢、需要认知努力的特点。两种加工过程相互竞争，获胜一方通过行为表达出来。与此同时，Haidt (2001, 2007)也提出了一个道德判断的社会直觉主义模型(social intuitionist model)。该模型的核心观点认为，道德判断是一种主要由情绪启动，具有快速、自动化特点的直觉反应过程。道德直觉即道德情境下的直觉反应，有意识的推理过程发生在直觉反应之后，作用是对道德直觉寻找合理的解释(田学红 等, 2011; 唐江伟 等, 2015; 徐平, 迟毓凯, 2007)。

在道德困境中，哪些是出于直觉反应，哪些是出于推理反应？研究发现，被试做功利主义判断所用的时间比作道义论判断更长(Greene et al., 2001)。相比于有时间压力，鼓励被试费力思考会做出更多功利主义的判断(Bartels, 2008; Suter & Hertwig, 2011)。增加认知负荷会使被试做功利主义判断的时间增长，但对道义论判断的时间没有影响(Greene et al., 2008)。被试经历了被直觉所误导的操纵后，更倾向于作功利主义的判断(Paxton et al., 2012; Pinillos et al., 2011)。这些研究结果表明，直觉反应导致了道义论的决择，推理反应导致了功利主义的决择。拒绝把人推下天桥的决定出自直觉反应，而不是出自经成本收益计算后的推理反应。

## 3 道德直觉的主观性

### 3.1 文化因素影响道德判断

早期以皮亚杰(Piaget)与科尔伯格(Kohlberg)为代表的道德心理学家认为，道德判断是由理性作出的(喻丰 等, 2011)。但Shweder等人(1987)的研究却发现，文化不同，对道德的理解也不同。他们收集了几十个违背社会规范的行为作为事例，对美国与印度的被试进行采访。结果发现，一些美国人认为是错误的行为，但印度人却认为是可以接受的(如儿子比女儿继承更多的财产)。也有一些印度人认为错误的行为，美国人却认为可以接受(如成年儿子对父亲直呼其名)。研究者由此认为，

在社会中心主义文化(如印度)中，道德是社会中心，个人之间相互依赖，道德原则(用以禁止伤害)和社会惯例(用以规范行为，与伤害不直接相关)之间没有明确的界限，但是，个人主义文化(如美国)中，许多社会惯例都没有纳入到道德原则中。

Haidt 等人(1993)采用了相似的研究模式探讨文化与道德的关系。他们以伤害、恶心和不敬为主旨编制了一系列事例，对来自不同文化地区的被试进行采访研究。结果发现，对于那些不涉及伤害但属于禁忌的行为(如烹食意外死亡的宠物、用国旗擦拭马桶等)，个人主义文化地区(美国)的被试倾向于认为是违反社会惯例，而社会中心主义地区(巴西)的被试则倾向于认为是对道德的侵犯。并且，当控制了“伤害”因素后，两者的差异更加凸显。这些结果反映出不同文化中道德的范围是不一样的。还有一些更为重要的发现是，约38%的受访对象听到的是没有伤害却具有冒犯的故事，但在回忆中却声称故事中有人受到了伤害。当要求被试解释所作道德判断的原因时，被试产生一种“道德之愕” (morally dumbfounded)的现象，即无法解释作出判断的理由。这些研究表明，人们凭借直觉快速做出道德判断，并且，同一类事件在不同文化背景下，人们所作出的道德判断也可能并不相同。

### 3.2　个人力量对道德直觉的影响

Greene 等人(2004)对“电车问题”的研究表明，人们在人行天桥困境与开关困境所做的判断大不相同，且人行天桥困境激发了人们更多的情感反应。那么，人行天桥困境的哪些因素触发了大脑的情感开关？研究者首先探讨了空间距离对道德两难选择的影响。为此，他们设计了一个远距离人行天桥困境(the remote footbridge dilemma)，该困境与人行天桥困境的不同之处在于，实施者不用亲手把人推下桥，而是站在远处，通过启动一个活板门开关，使人掉下铁轨并阻挡电车。结果表明，赞同“以一换五”的比例显著高于原始人行天桥困境(Greene et al., 2009)。远距离人行天桥困境中，实施者距离受害者远，而且没有直接触碰受害者。那么，是距离还是触碰，或者两者共同对道德判断起作用？为了解答这一问题，研究者设计了人行天桥开关困境(The footbridge switch dilemma)，该困境与远距离人行天桥困境的不同之处在于，活板门的开关不是在远处，而是在受害者旁边。结果表明，赞同“以一换五”的比例与远距离人行天桥困境无显著差异(Greene et al., 2009)。因此，空间距离并非影响人们作出道德选择的关键因素。

在人行天桥困境中，实施者直接触碰受害者。人行天桥开关困境则没有发生直接触碰，因此，触碰可能是一个关键因素。但人行天桥困境除了触碰外，还直接对受害者施加力量，即使用了个人力量。为了区分触碰与个人力量，研究者设计了人行天桥长杆困境(The footbridge pole dilemma)，该困境与人行天桥困境的不同之处在于，实施者通过一个长杆推动受害者，创造了一个不发生直接触碰，却运用了个人力量的情境。结果表明，赞同“以一换五”的比例与人行天桥困境没有显著差异。因此，个人力量的运用是导致被试在人行天桥困境与开关困境决策差异的最主要因素。人行天桥困境中，运用个人力量伤害他人触发了大脑的情感开关(Moore et al., 2008; Young et al., 2006)。

道德直觉表现出了不灵敏的一面，让人从活门掉下去也许是不正确的，但因为没有使用个人力量“亲自推”，就没有引起足够的警惕。道德直觉也表现出了过于灵敏的一面，与“不推”导致的五倍伤害相比，“推”可能是正确的选择，直觉反应对受害人承受的伤害过于担心。从这个角度上看，道德直觉并不十分可靠。

### 3.3　手段与连带作用对道德直觉的影响

开关困境与人行天桥困境还有一个重要的区别：开关困境中对工人的伤害是一种连带作用(若工人此时突然走开，则能避免伤害)。而人行天桥困境中的伤害是把工人的身体作为阻挡电车的手段(若工人此时突然走开，计划便失败)。道德直觉对手段与连带作用的区分敏感吗？为了解答这一问题，研究者设计了障碍物冲撞困境(The obstacle collide dilemma)。该困境与开关困境类似，不同之处在于实施者需要跑过一条狭窄的通道才能到达开关处，通道上站着一个人，实施者跑过通道时必然会把此人撞下天桥摔死。在这场景中，实施者是通过个人力量将人撞下桥摔死的，但伤害是由连带作用导致的。结果显示，障碍物冲撞困境中，大多数人赞同“以一换五”，与开关困境的赞同率没有显著差异(Cushman et al., 2006; Greene et al., 2009; Schaich Borg et al., 2006)，表明道德

直觉对手段与连带作用的区分是敏感的。但是，在远距离人行天桥困境和人行天桥开关困境中，都把人的身体作为阻挡电车的手段，为何赞同“以一换五”的比例却远远高于人行天桥困境？为解释这一问题，研究者引入了环线铁路困境(The loop case)进行探讨，该困境与开关困境类似，不同之处在于支线铁路与主线铁路是相通的，如果支线铁路上没有人，电车会回到主线铁路，撞向五名工人。在该场景中，没有使用个人力量，但把人的身体作为阻挡电车的手段(若工人消失，计划便失败)。结果发现，“以一换五”的赞同率与开关困境没有显著差异(Greene et al., 2009; Thomson, 1985)。

环线铁路困境的特点是能明显判断没有使用个人力量，但却不容易判断伤害是手段还是连带作用引致(Bruers & Braeckman, 2014; Kaufman, 2016)。尽管该困境较为复杂，但它与情感、认知均紧密相关(Conway et al., 2018; Kahane et al., 2017)。环线铁路困境的结果表明，若“伤害是达成目标的手段”以一种较为复杂的形式呈现时，人们可能会接受伤害，即赞同利用人的身体阻挡电车。

Greene (2013)在总结以上一系列的电车困境后发现，个人力量、手段/连带作用是影响人们直觉判断的两个重要因素，但这两个因素结合起来才会产生稳定的影响。使用个人力量，但伤害只是连带作用，这种行为可接受程度较高(如障碍物冲撞困境)。如果伤害是达成目标的手段，但没有使用个人力量，这种行为的可接受程度同样较高(如环线铁路困境)。如果伤害是达成目标的手段，并且使用了个人力量，这种行为的可接受程度较低(如人行天桥困境与人行天桥长杆困境)。

## 4 “模块近视假说”的解释

### 4.1 决策中的双重加工过程

推理决策领域中的双加工理论(the Dual-Process Theory)最初由 Evans 和 Wason (1976)提出，并得到不断的修正与发展，是该领域中颇有影响力的理论(胡竹菁，胡笑羽，2012；肖前国 等，2009)。双加工理论把人的认知加工过程区分成类型 1 (T1)和类型 2 (T2)。由于双加工理论在不同发展时期及不同学者中的内涵、特征并不完全一致，并由此遭受批评。Evans 等人在回应中指出，区分不同加工的核心特征是工作记忆资源的参与度(Evans & Stanovich, 2013；艾炎，胡竹菁, 2018)。综合地说，T1 实行并行加工，具有占用心理资源少，反应自动化，但反应方式刻板的特点。T2 实行串行加工，占用心理资源较多，且加工速度较慢，但不易受刻板印象的影响，能作出相对灵活的反应。两个加工过程均对决策起重要作用，根据两者作用方向的一致性，呈现协同或竞争关系(孙彦 等, 2007)。

研究者对这两种加工类型在时间进程上持不同的观点，产生了三类有代表性的模型(艾炎，胡竹菁，2018)。序列加工模型(Serial processing model)认为，T1 首先开始加工信息，并产生默认反应，T2 加工根据时间压力、个人认知倾向等因素决定是否介入并产生替代反应(Evans, 2011; Evans & Stanovich, 2013)。第二类是平行竞争模型(Parallel competition model)，该模型认为两个加工系统同时启动，并行运作。T1 加工的速度更快，但需要等加工速度慢的 T2 加工完成后才能做出正确判断。该模型得到了一些实验结果的支持(Epstein, 1994; Handley et al., 2011; Handley & Trippas, 2015; Sloman, 1996; Stupple & Ball, 2008; Trippas et al., 2016; Trippas et al., 2017)。第三类是混合模型(Hybrid model)，它与前面两种模型的最大区别在于认为冲突探查是一个重要且独立的过程，并且是决定 T2 加工是否介入的重要因素。混合模型的主要代表之一是分析参与的三阶段双重加工模型(Three-stage dual-process model of analytic engagement)。该模型认为，在 T1 初始反应阶段，自动加工可能产生多种直觉反应。第二阶段是对 T1 中潜在的冲突探测，若没有探测到冲突，则继续进入第三阶段，T2 粗略分析并接受这一反应。但若探测到冲突，T2 产生分析性加工。T2 有可能产生致力于证实初始反应的分析，即合理化(rationalize)，也有可能对初始反应进行抑制，产生替代性反应，即解耦(decouple) (Pennycook et al., 2015; Pennycook, 2017)。

### 4.2 “近视”的监控系统

研究者在决策与推理研究的双重加工理论框架下，探索人们道德判断的认知加工机制。Haidt (2001, 2007)质疑传统道德理性推理模型忽略了直觉加工，并提出了一个道德判断的社会直觉主义模型，把直觉加工纳入考虑。道德直觉在道德判断中的作用得到了许多研究者的重视，但在对

上述研究的梳理中可以发现, 道德直觉具有强烈的主观性。在“电车问题”中, 道德直觉对运用个人力量且有意造成的伤害行为高度警惕, 对没有运用个人力量或由连带作用产生的伤害行为却不敏感。

Greene 在双加工理论的框架下提出了一个“模块近视假说” (modular myopia hypothesis), 用以解释个人力量、手段与连带作用对道德直觉的作用。该假说认为, 人类大脑包含一个认知子系统(模块), 能自动化地监控审查个体的计划行为, 并在个体打算伤害他人时发出情绪性的警报。但是, 该系统只调用少量的认知资源参与, 是一种简单的“单通道”审查加工, 无法对造成伤害的多条因果链同时监控。因此, 当人们有意图地把“伤害他人作为达成目标的手段”时, 即伤害计划位于主因果链上, 认知子系统中的监控机制能察觉并发出情绪警报, 警惕伤害行为。而如果“伤害”是由连带作用引发, 即“伤害”位于次级因果链上时, 由于监控审查系统的认知资源局限, “连带伤害”避开了审查机制的监控, 使其不能引发情绪警报, 即该系统是“近视眼” (myopic), 无法识别出由连带作用造成的伤害(Greene, 2013)。

“模块近视假说”可以纳入“分析参与的三阶段双重加工模型”中加以理解。在道德判断中的第一阶段, 自动化加工的 T1 依赖于信息特征产生初始计划反应(如在“电车问题”中依据数量关系产生接受“以一换五”的计划), 第二阶段, 认知子系统(模块)对个体计划行为进行探查, 对有意图的伤害行为发出情绪性警报。但该模块进行的是浅层次的加工, 只调用少量的认知资源, 具有刻板性。如果模块没有发出情绪警报, 初始计划反应继续进入第三阶段, T2 加工粗略分析并接受该反应(即接受“以一换五”)。但是, 如果模块发出了情绪警报, 则拒绝初始计划反应(即拒绝“以一换五”)。在充足的时间资源下, T2 的分析性加工介入, 如果 T2 对初始计划反应的成本效益分析产生足够的说服力, T2 加工也可能压制情绪警报, 坚持初始计划反应。

“模块近视假说”合理地解释了为什么在环线铁路困境中, 即便把伤害作为达成目标的手段, 这种行为的可接受程度同样较高。环线铁路困境中, 扳动开关是利用受害者阻挡电车, 即把伤害作为达成目标的手段。但该困境复杂程度较高, 需要通过里面的多条因果链, 才能判断出对受害者的伤害是达成目标的手段。在环线铁路困境的主因果链是: 若电车沿直线行驶, 则会撞死 5 名工人。扳动转向开关就破坏了主因果链, 但是, 因为环线铁路困境中电车的支线是与主线相连的。因此, 电车转向后, 又形成了次级因果链: 若不用一个人的身体挡住电车, 电车会继续撞向 5 名工人, 即“把伤害作为达成目标的手段”发生在次级因果链上。由于审查机制的“单通道”局限, 次级因果链的有害行为不能被监控系统识别, 因此也不会发出情绪警报(Greene, 2013)。“模块近视假说”解释了“电车问题”中道德直觉的主观性与刻板性。

“模块近视假说”表明, 在相似的道德情境中, 看似无关道德的因素有可能成为触发人们大脑情感警报的关键, 导致人们产生截然不同的判断, 造成了道德直觉的不可靠。这种现象得到了许多研究证据的支持, Schelling (1968)发现, 与不确定或以数字为代表的个体相比, 人们对可辨识个体的需求会做出更加迅速、急切的反应。特定个体受到的伤害更能唤起人们的情绪, 这种现象被称为“可识别受害者效应” (identifiable victim effect)。在 Small 和 Loewenstein (2003)的研究中也发现, 在捐助情境中, 人们对“可识别”对象的同情与慷慨程度高于“以数字代表”的对象。而且, 如果捐赠的对象是某个具体需要帮助的人, 人们会有较强的捐赠意愿, 但如果是为了类似于解决非洲的贫困这类范围更大的问题时, 人们的捐赠意愿会减弱。Kogut 和 Ritov (2005)的研究中, 设置了两种捐助情境, 一种是 1 名需要巨额医疗费用的患病儿童, 另一种是 8 名类似情况的儿童, 结果发现, 人们对 1 名儿童表现出更强的同情心。从理性上分析, 受助方是否可辨识不应该是影响人们道德行为的因素, 但是, 由于道德直觉的刻板性, 人们会不自觉地对触发情绪的情境给予更多的道德关怀。

## 5　总结与展望

研究者们采用道德两难困境为研究范式, 从行为学、认知神经科学及神经病理学层面探讨了道德判断的认知机制。大量的研究结果表明, 人们主要依靠直觉做出道义论的判断, 情感是其中的决定性因素。但是, 面对相似的道德情境, 道德直觉表现出了不可靠性。Greene 提出了一个“模块

近视假说”对此进行了解释。该假说认为，由直觉系统发出的情绪性警报会阻止人们作出伤害的行为。但是，由于直觉系统“单通道”式的加工局限，使得由非个人力量或连带作用造成的伤害不能触发情绪警报，造成了道德直觉在相似道德情境中的主观性。

“模块近视假说”对理解现实生活中的道德问题提供了有益的帮助。功利主义对更大利益的公正关注具有积极、利他的一面(Kahane et al., 2017)，但道德直觉对情绪性警报可能反应过度，拒绝那些会引起强烈情绪反应的功利主义决策。如在生物、医学领域中，动物试验、安乐死、堕胎等伪暴力行为往往会促进更大范围的利益，但人们可能会出于情感上的厌恶给予拒绝。在 Greene 等人的研究中发现，即使把人行天桥类似困境中的伤害增加上百万倍，赞同把人推下天桥的比率依然低于开关困境(Greene et al., 2009; Paxton et al., 2012)。直觉反应对情绪性警报也可能会过于迟钝，忽视那些没有引起强烈情绪反应的事件的负面影响。如在环境污染、全球气候变暖等问题上，我们对他人造成伤害时，伤害的方式几乎总是以连带作用的方式出现，也没有直接使用个人力量。

随着自动驾驶汽车的应用，“电车问题”逐渐从思维实验演变成了现实生活中真实面对的问题，人们对“电车问题”的研究模式与研究结果有望为探讨如何应对自动驾驶中的道德困境提供参考。在自动驾驶汽车困境的研究中发现，在虚拟现实环境中，作为司机体验的参与者在虚拟驾驶中更多使用功利主义决策，这一结论可以为自动驾驶中的决策算法提供指导(Faulhaber et al., 2018)。但是，也有研究者认为，在不同文化环境中，由于价值观、道德观念、社会规范、道路规则以及政策等方面的差异，人们对自动驾驶道德规范的态度也会有所不同(Greene, 2016b; Taddeo & Floridi, 2018)。仅依靠功利主义的自动驾驶观念是一种道德一元论，并不适合多元化的社会。为此，一些研究人员提出了不同的自动驾驶道德伦理观，试图建立多元性的自动驾驶道德规范(Bergmann et al., 2018; Contissa et al., 2017)。最近有研究者把文化维度纳入了自动驾驶伦理框架中，他们比较了集体主义(韩国)和个人主义(加拿大)两种不同文化下人们的道德推理，在人们身上印证了利他主义道德论、道德非决定论和道德义务论这三种道德推理类型。这种研究模式降低了自动驾驶道德问题研究的抽象性，为特定文化的设计提供了指导(Rhim et al., 2020)。在个体层面，人们面对自动驾驶道德问题时，同样表现出主观、甚至矛盾的一面，人们一方面赞同使用“为了更大的利益可以牺牲乘客”的自动驾驶软件。但另一方面，他们自己更愿意乘坐不惜一切代价保护乘客的汽车(Bonnefon et al., 2016; Lawless & Postnikov, 2016; Maxmen, 2018)。人们在道德判断中的不客观提示我们，面对道德争议，不应仅仅听凭道德直觉，深思熟虑是一种同样极具价值的行动依据。

## 参考文献


艾炎，胡竹菁. (2018). 推理判断中双重加工过程的协作与转换机制. *心理科学进展, 26*(10), 1794–1806.

胡竹菁，胡笑羽. (2012). Evans 双重加工理论的发展过程简要述评. *心理学探新, 32*(4), 310–316.

孙彦，李纾，殷晓莉. (2007). 决策与推理的双系统——启发式系统和分析系统. *心理科学进展, 15*(5), 721–726.

唐江伟，路红，刘毅，彭坚. (2015). 道德直觉决策及其机制探析. *心理科学进展, 23*(10), 1830–1842.

田学红，杨群，张德玄，张烨. (2011). 道德直觉加工机制的理论构想. *心理科学进展, 19*(10), 1426–1433.

肖前国，罗乐，余林. (2009). 推理与决策的双加工理论研究简评. *心理科学进展, 17*(2), 321–324.

徐平，迟毓凯. (2007). 道德判断的社会直觉模型述评. *心理科学, 30*(2), 403–405.

喻丰，彭凯平，韩婷婷，柴方圆，柏阳. (2011). 道德困境之困境——情与理的辩争. *心理科学进展, 19*(11), 1702–1712.

Awad, E., Dsouza, S., Shariff, A., Rahwan, I., & Bonnefon, J-F. (2019). Universals and variations in moral decisions made in 42 countries by 70, 000 participants. *Proceedings of the National Academy of Sciences, 117*(5), 2332–2337.

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition, 108*(2), 381–417.

Bergmann, L. T., Schlicht, L., Meixner, C., König, P., Pipa, G., Boshammer, S., Stephan, A. (2018). Autonomous vehicles require socio-political acceptance—an empirical and philosophical perspective on the problem of moral decision making. *Frontiers in Behavioral Neuroscience, 12*, 31.

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science, 352*(6293), 1573–1576.

Bruers, S., & Braeckman, J. (2014). A review and systematization of the trolley problem. *Philosophia, 42*(2), 251–269.

Contissa, G., Lagioia, F., & Sartor, G. (2017). The ethical knob: Ethically-customisable automated vehicles and the law. *Artificial Intelligence and Law, 25*, 365–378.

Conway, P., Goldstein-Greenwood, J., Polacek, D., & Greene,

J. D. (2018). Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition, 179,* 241–265.

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences, 107*(40), 17433–17438.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*(12), 1082–1089.

Elisa, C., Michela, M., Elisabetta. L., & Giuseppe, D. P. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience, 2*(2), 84–92.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*(8), 709–724.

Evans, J. S. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review, 31*(2-3), 86–102.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223–241.

Evans, J. S. B. T., & Wason, P. C. (1976). Rationalization in a reasoning task. *British Journal of Psychology, 63*(4), 479–486.

Faulhaber, A. K., Dittmer, A., Blind, F., WäChter, M. A., Timm, S., Sütfeld, L. R., … König, P. (2018). Human decisions in moral dilemmas are largely described by utilitarianism: Virtual car driving study provides guidelines for autonomous driving vehicles. *Science and Engineering Ethics, 25*(2), 399–418.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review, 2*(2), 152–161.

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter. M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of Personality and Social Psychology, 113*(3), 343–376.

Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R., & Hütter. M. (2018). Effects of incidental emotions on moral dilemma judgments: An analysis using the CNI model. *Emotion, 18*(7), 989–1008.

Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. New York: Penguin Press, 105–132.

Greene, J. D. (2016a). Solving the Trolley Problem. *A Companion to Experimental Philosophy*, John Wiley & Sons, Ltd, 175–189.

Greene, J. D. (2016b). Our driverless dilemma. *Science, 352*(6293), 1514–1515.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111*(3), 364–371.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107*(3), 1144–1154.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834.

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*(5827), 998–1002.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*(4), 613–628.

Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 28–43.

Handley, S. J., & Trippas, D. (2015). Dual processes and the interplay between knowledge and structure: A new parallel processing model. *Psychology of Learning and Motivation, 62*, 33–58.

Kahane, G., Everett, J. A. C., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2017). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review, 125*(2), 131–164.

Kaufman, W. R. P. (2016). The doctrine of double effect and the trolley problem. *Journal of Value Inquiry, 50*(1), 21–31.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature, 446*(7138), 908–911.

Kogut, T., & Ritov, I. (2005). The "Identified Victim" Effect: An Identified Group, or Just a Single Individual? *Journal of Behavioral Decision Making, 18*(3), 157–167.

Lawless, W. F., & Postnikov, V. (2016). Autonomous vehicles. *Science, 354*(6311), 426.

Maxmen, A. (2018). Self-driving car dilemmas reveal that moral choices are not universal. *Nature, 562*(7728), 469–470.

Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgement in Frontotemporal dementia. *Cognitive and Behavioral Neurology, 18*(4), 193–197.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? individual differences in working memory capacity, executive control, and moral judgment. *Psychological*

*Science, 19*(6), 549–557.
Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science, 36*(1), 163–177.
Pennycook, G. (2017). *A Perspective on the Theoretical Foundation of Dual Process Models*. In W. De Neys (Ed.) *Dual Process Theory 2.0.* New York, NY: Psychology Press.
Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? a three-stage dual-process model of analytic engagement. *Cognitive Psychology, 80,* 34–72.
Perkins, A. M., Leonard, A. M., Weaver, K., Dalton, J. A., Mehta, M. A., Kumari, V., … Ettinger, U. (2013). A dose of ruthlessness: Interpersonal moral judgment is hardened by the anti-anxiety drug lorazepam. *Journal of Experimental Psychology: General, 142*(3), 612–620.
Pinillos, N. Á., Smith, N., Nair, G. S., Marchetto, P., & Mun, C. (2011). Philosophy's new challenge: Experiments and intentional action. *Mind & Language, 26*(1), 115–139.
Rhim, J., Lee, G. B., & Lee, J. H. (2020). Human moral reasoning types in autonomous vehicle moral dilemma: A cross-cultural comparison of korea and canada. *Computers in Human Behavior, 102*, 39–56.
Schaich Borg, J., Hynes, C., van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience, 18*(5), 803–817.
Schelling, T. C. (1986). The life you save may be your own. In S. B. Chase (Ed), *Problesms in public expenditure analysis* (pp.127−162). Washington DC: The Brookings Institute.
Shweder, R. A., Mahapatra, M, & Miller, J. G. (1987). Culture and Moral Development. In J. Kagan, & S. Lamb (Eds.), *The Emergence of Moral in Young Children* (pp. 1–83). Chicago, American: University of Chicago Press.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3–22.
Small, D. A., & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty, 26,* 5–16.
Stupple, E. J. N., & Ball, L. J. (2008). Belief–logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning, 14*(2), 168–181.
Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition, 119*(3), 454–458.
Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science, 361*(6404), 751–752.
Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist, 59*(2), 204–217.
Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal, 94*(6), 1395–1415.
Trémolière, B., de Neys, W., & Bonnefon, J. F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition, 124*(3), 379–384.
Trippas, D., Handley, S. J., Verde, M. F., & Morsanyi, K. (2016). Logic brightens my day: Evidence for implicit sensitivity to logical validity. *Journal of Experimental Psychology Learning Memory and Cognition, 42*(9), 1448–1457.
Trippas, D., Thompson, V. A., & Handley, S. J. (2017). When fast logic meets slow belief: Evidence for a parallel processing model of belief bias. *Memory & Cognition, 45*(4), 539–552.
Young, L., Cushman, F., Adolphs, R., Tranel, D., & Hauser, M. (2006). Does emotion mediate the relationship between an action's moral status and its intentional status? neuropsychological evidence. *Journal of Cognition and Culture, 6*(1-2), 291–304.
Young, L., & Koenigs, M. (2007). Investigating emotion in moral cognition: A review of evidence from functional neuroimaging and neuropsychology. *British Medical Bulletin, 84*(1), 69–79.

## Moral intuition is moral but not objective

YUAN Xiaojing, LIU Chang
(*School of Psychology, Nanjing Normal University, Nanjing* 210024, *China*)

**Abstract:** Previous studies have shown that certain moral intuitions prompt people to make deontological judgments. However, moral intuition, dominated by emotional factors, is relatively subjective. Highlights intentional harm through personal power, moral intuition ignores harm caused by non-personal forces and side effects. The Modular Myopia Hypothesis states a cognitive subsystem present in the human brain, monitoring an individual's planned behavior and releasing emotional alerts when it detects a harmful event in an action plan. But this inspector is a relatively simple, "single-channel" system that cannot monitor multiple causal chains simultaneously and fails to identify harmful side effects. The subjectivity of moral intuition suggests that in the face of moral disputes in real life, people should not rely on intuition as the sole basis for decisions and actions.

**Key words:** moral dilemma, moral intuition, emotions, modular myopia hypothesis